```
RESULT 1
TMS5 HUMAN
                                  PRT;
ID
    TMS5 HUMAN
                     TANDARD;
     Q9H3S3;
AC
     16-OCT-2001 (Rel. 40, Created)
DT
     16-OCT-2001 (Rel. 40, Last sequence update)
DT
    28-FEB-2003 (Rel. 41, Last annotation update)
DT
    Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
DE
    TMPRSS5.
GN
    Homo sapiens (Human).
OS
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC
OX
    NCBI TaxID=9606;
RN
     [1]
    SEQUENCE FROM N.A.
RΡ
RC
    TISSUE=Brain;
    PubMed=11741986;
RX
RA
    Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
     "Spinesin/TMPRSS5, a novel transmembrane serine protease, cloned from
RT
    human spinal cord.";
RT
    J. Biol. Chem. 277:6806-6812(2002).
RL
CC
     -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
     -!- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
CC
        neurons, in their axons, and at the synapses of motoneurons in the
CC
CC
        spinal cord.
     -!- SIMILARITY: Belongs to peptidase family S1.
CC
CC
     -!- SIMILARITY: Contains 1 SRCR domain.
     ______
CC
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
CC
DR
    EMBL; AB028140; BAB20375.1; -.
DR
    HSSP; P00763; 1DPO.
DR
    Genew; HGNC:14908; TMPRSS5.
    MIM; 606751; -.
DR
DR
    MEROPS; S01.313; -.
DR
    InterPro; IPR009003; Cys Ser trypsin.
    InterPro; IPR001254; Peptidase S1.
DR
    InterPro; IPR001314; Peptidase S1A.
DR
     InterPro; IPR001190; Srcr receptor.
DR
DR
     Pfam; PF00089; trypsin; 1.
DR
     PRINTS; PR00722; CHYMOTRYPSIN.
DR
    SMART; SM00020; Tryp SPc; 1.
DR
     PROSITE; PS50240; TRYPSIN DOM; 1.
     PROSITE; PS00134; TRYPSIN HIS; 1.
     PROSITE; PS00135; TRYPSIN SER; 1.
DR
DR
     PROSITE; PS00420; SRCR 1; FALSE NEG.
     PROSITE; PS50287; SRCR 2; FALSE NEG.
DR
KW
    Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW
    Glycoprotein.
FT
    DOMAIN
                        49
                                 CYTOPLASMIC (POTENTIAL).
                  1
FT
     TRANSMEM
                 50
                        70
                                 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
```

```
(POTENTIAL).
FT
                             EXTRACELLULAR (POTENTIAL).
               71
FT
    DOMAIN
                    457
    DOMAIN
                    207
                             SRCR.
FT
              112
                             SERINE PROTEASE.
FT
    DOMAIN
              218
                    457
                             CHARGE RELAY SYSTEM (BY SIMILARITY).
    ACT SITE
FT
              258
                    258
                             CHARGE RELAY SYSTEM (BY SIMILARITY).
FT
    ACT SITE
              308
                    308
    ACT SITE
                             CHARGE RELAY SYSTEM (BY SIMILARITY).
FT
              405
                    405
    SITE
                             CLEAVAGE (POTENTIAL).
FT
              217
                    218
                             BY SIMILARITY.
TT
    DISULFID
              135
                    196
    DISULFID
              148
                    206
                             BY SIMILARITY.
FT
FT
    DISULFID
              209
                    328
                             BY SIMILARITY.
                    259
FT
    DISULFID
              243
                             BY SIMILARITY.
FT
    DISULFID
              374
                    390
                             BY SIMILARITY.
                    429
                             BY SIMILARITY.
FT
    DISULFID
              401
                             N-LINKED (GLCNAC. . .) (POTENTIAL).
    CARBOHYD
              163
                    163
FT
                             N-LINKED (GLCNAC. . .) (POTENTIAL).
              170
                    170
FT
    CARBOHYD
                             N-LINKED (GLCNAC. . .) (POTENTIAL) .
    CARBOHYD
              195
                    195
FT
                             N-LINKED (GLCNAC. . .) (POTENTIAL).
FT
    CARBOHYD
              319
                    319
                             N-LINKED (GLCNAC. . .) (POTENTIAL).
FT
    CARBOHYD
              375
                    375
                              64406AB4985A2651 CRC64;
SQ
    SEQUENCE
             457 AA;
                     49574 MW;
                      100.0%;
                             Score 2481; DB 1;
                                               Length 457;
 Query Match
                      100.0%; Pred. No. 5.6e-194;
 Best Local Similarity
 Matches 457; Conservative
                            0; Mismatches
                                           0;
                                               Indels
                                                       0;
                                                           Gaps
                                                                  0;
          1 MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPISQAVCWRSMRRGCAVLGALGLLAG 60
Qу
            1 MSLMLDDQPPMEAQYAEEGPGPGIFRAEPGDQQHPISQAVCWRSMRRGCAVLGALGLLAG 60
Db
         61 AGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEALLPALPKTVSFRINSEDFL 120
Qу
            Db
         61 AGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEALLPALPKTVSFRINSEDFL 120
        121 LEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVNLTDIKLNSSQEFAQLSP 180
Qy
            121 LEAOVRDOPRWLLVCHEGWSPALGLOICWSLGHLRLTHHKGVNLTDIKLNSSQEFAQLSP 180
Db
        181 RLGGFLEEAWQPRNNCTSGQVVSLRCSECGARPLASRIVGGQSVAPGRWPWQASVALGFR 240
Qу
            181 RLGGFLEEAWQPRNNCTSGQVVSLRCSECGARPLASRIVGGQSVAPGRWPWQASVALGFR 240
Db
        241 HTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHQGALVERIIPHPLY 300
Qу
            241 HTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHSAVRPHOGALVERIIPHPLY 300
Db
Qy
        301 SAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHFPKGSRCWVSGWGHTHPSHTYSSDM 360
            301 SAONHDYDVALLRLOTALNFSDTVGAVCLPAKEOHFPKGSRCWVSGWGHTHPSHTYSSDM 360
Db
Qу
        361 LQDTVVPLFSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLV 420
            361 LQDTVVPLFSTQLCNSSCVYSGALTPRMLCAGYLDGRADACQGDSGGPLVCPDGDTWRLV 420
Db
        421 GVVSWGRACAEPNHPGVYAKVAEFLDWIHDTAQDSLL 457
Qy
            421 GVVSWGRACAEPNHPGVYAKVAEFLDWIHDTAQDSLL 457
Db
```

```
RESULT 1
S00845
hepsin (EC 3.4.21.-) - human
C; Species: Homo sapiens (man)
C;Date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text change 18-Jun-1999
C; Accession: S00845
R; Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A; Title: A novel trypsin-like serine protease (hepsin) with a putative
transmembrane domain expressed by human liver and hepatoma cells.
A; Reference number: S00845; MUID: 88209431; PMID: 2835076
A; Accession: S00845
A; Molecule type: mRNA
A; Residues: 1-417 <LEY>
A;Cross-references: EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
C; Genetics:
A; Gene: GDB: HPN; TMPRSS1; hepsin
A; Cross-references: GDB:135685; OMIM:142440
A; Map position: 19q11-19q13.2
C; Superfamily: hepsin; trypsin homology
C; Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F;23-45/Domain: transmembrane #status predicted <TMN>
F;163-400/Domain: trypsin homology <TRY>
F;188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F;203,257,353/Active site: His, Asp, Ser #status predicted
                       24.3%; Score 602.5; DB 1; Length 417;
  Query Match
 Best Local Similarity 33.3%; Pred. No. 2.9e-41;
 Matches 139; Conservative 61; Mismatches 155; Indels
          53 GALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEASAEEALLPALPKTVSF 112
Qy
                                                  : |: :| |
             24 GTLLLLTAIGAASWAIVAVLLRSDQEPLYPV-----QVSSADARLMVFDKT--- 69
Dh
         113 RINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRLTHHKGVNLTDIKLNSS 172
Qy
                                    : : | :| || | :::
                          : | |:|
          70 -----EGTWRLLCSSRSNARVAGLSCEEMGFLRALTHSELDVRTAGANGT 114
Db
         173 QEFAQLSPRLGGFLEEAWQPRN------NCTSGQVVSLRCSECGARPL-ASRIVGG 221
Qy
                                         ::| |
         115 SGFF-----CVDEGRLPHTQRLLEVISVCDCPRGRFLAAICQDCGRRKLPVDRIVGG 166
Db
         222 QSVAPGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAGLVSHS 281
QУ
                             : ||||||
                                                _|| || || || |: :
         167 RDTSLGRWPWQVSLRYDGAHLCGGSLLSGDWVLTAAHCFPE-RNRVLSRWRVFAGAVAQA 225
Db
         282 AVRPHOGAL-VERIIPHPLY-----SAONHDYDVALLRLQTALNFSDTVGAVCLPAKEQ 334
QУ
                   226 S--PHGLQLGVQAVVYHGGYLPFRDPNSEENSNDIALVHLSSPLPLTEYIQPVCLPAAGQ 283
Db
         335 HFPKGSRCWVSGWGHTHPSHTYSSDMLODTVVPLFSTOLCNSSCVYSGALTPRMLCAGYL 394
Qу
               284 ALVDGKICTVTGWGNTQ-YYGQQAGVLQEARVPIISNDVCNGADFYGNQIKPKMFCAGYP 342
Db
         395 DGRADACOGDSGGPLVCPDG----DTWRLVGVVSWGRACAEPNHPGVYAKVAEFLDWI 448
Qу
                                    1||| ||::| :||
         343 EGGIDACQGDSGGPFVCEDSISRTPRWRLCGIVSWGTGCALAQKPGVYTKVSDFREWI 400
Db
```

```
RESULT 5
US-10-607-565-108
 Sequence 108, Application US/10607565
 Publication No. US20040048294A1
 GENERAL INFORMATION:
  APPLICANT: Rosen et \al.
  TITLE OF INVENTION: 31 Human Secreted Proteins
  FILE REFERENCE: PZ033P1
  CURRENT APPLICATION NUMBER: US/10/607,565
  CURRENT FILING DATE: 2003-06-27
  PRIOR APPLICATION NUMBER: US/09/531,119
  PRIOR FILING DATE: 2000-03-20
  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/101,546
  PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-23
  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/102,895
  PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-02
  NUMBER OF SEQ ID NOS: 140
  SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 108
   LENGTH: 480
   TYPE: PRT
   ORGANISM: Homo sapiens
US-10-607-565-108
 Query Match
                     91.9%; Score 2280; DB 12; Length 480;
 Best Local Similarity 99.8%; Pred. No. 2.9e-198;
 Matches 420; Conservative
                           0; Mismatches
                                             Indels
                                                     0; Gaps
         37 SQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEA 96
Qу
           Db
         60 SQAVCWRSMRRGCAVLGALGLLAGAGVGSWLLVLYLCPAASQPISGTLQDEEITLSCSEA 119
         97 SAEEALLPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRL 156
Qу
           Db
        120 SAEEALLPALPKTVSFRINSEDFLLEAQVRDQPRWLLVCHEGWSPALGLQICWSLGHLRL 179
        157 THHKGVNLTDIKLNSSQEFAQLSPRLGGFLEEAWQPRNNCTSGQVVSLRCSECGARPLAS 216
Qy
           Db
        180 THHKGVNLTDIKLNSSQEFAQLSPRLGGFLEEAWQPRNNCTSGQVVSLRCSECGARPLAS 239
        217 RIVGGQSVAPGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAG 276
Qу
           240 RIVGGQSVAPGRWPWQASVALGFRHTCGGSVLAPRWVVTAAHCMHSFRLARLSSWRVHAG 299
Db
Qy
        277 LVSHSAVRPHQGALVERIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEQHF 336
           300 LVSHSAVRPHQGALVERIIPHPLYSAQNHDYDVALLRLQTALNFSDTVGAVCLPAKEOHF 359
Db
Qy
        337 PKGSRCWVSGWGHTHPSHTYSSDMLQDTVVPLFSTQLCNSSCVYSGALTPRMLCAGYLDG 396
           Db
        360 PKGSRCWVSGWGHTHPSHTYSSDMLQDTVVPLFSTQLCNSSCVYSGALTPRMLCAGYLDG 419
Qy
        397 RADACQGDSGGPLVCPDGDTWRLVGVVSWGRACAEPNHPGVYAKVAEFLDWIHDTAQDSL 456
           Db
        420 RADACQGDSGGPLVCPDGDTWRLVGVVSWGRGCAEPNHPGVYAKVAEFLDWIHDTAODSL 479
        457 L 457
Qy
```